DOCUMENT RESUME

ED 261 095 TM 850 521

AUTHOR Patton, Michael Quinn
TITLE Six Honest Serving Men for Evaluation.
PUB DATE 85
NOTE 40p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Evaluation; *Evaluation Methods; Evaluation Needs; *Evaluation Utilization; Evaluators; *Program Evaluation
IDENTIFIERS Caribbean Agricultural Extension Project; Evaluation Problems; Evaluation Research

ABSTRACT

This paper reviews what has been learned about evaluation ultilization during the past 20 years. Evaluation utilization is discussed in terms of what is used, who uses evaluation, when evaluation is used, how evaluation is used, where evaluation is used, and why evaluation is used. It is suggested that the personal factor - the interests and commitments of key people involved in the evaluation - is the most important explanatory variable in evaluation utilization. Ten basic utilization-focused evaluation premises are presented, and it is suggested that the primary direction for further utilization research is to test these premises in an integrated, holistic way in real evaluations. The recent external evaluation of the Caribbean Agricultural Extention Project is described to illustrate a comprehensive test of utilization focused evaluation in a real world project. Five issues that continue to create confusion in the field and that emerge from discussions of the stakeholder assumption and research on the personal factor are discussed. It is concluded that evaluators should act on what is already known about utilization and evaluate those actions rather that do further isolated research on utilization. Preliminary observations on the problems of misutilization are also presented. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Patton, Michael Quinn
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED261095

SIX HONEST SERVING MEN FOR EVALUATION

By Michael Quinn Patton
University of Minnesota

The practice of evaluation involves the systematic collection of information about the activities, characteristics, and outcomes of programs, personnel, and products for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs, personnel, or products are doing and affecting. This definition of evaluation emphasizes (1) a systematic collection of information about (2) a broad range of topics (3) for use by specific people (4) for a variety of purposes.

This definition of evaluation is purposefully broad and includes, quite explicitly, the notion that evaluation can be used in a variety of ways. Definitions of evaluation vary, with some being quite narrow. In reviewing research on the utilization of evaluation, it is helpful to begin with how evaluation is defined because variations in definitions of evaluation lead to variations in definitions of utilization, and make the task of conducting research on utilization quite complex and varied.

Variations in Evaluation Definitions

William J. Gephart's comprehensive effort at defining evaluation illustrates both the problem and one kind of solution. He begins with the assumption that no short, succinct definition will suffice. Single-sentence definitions usually contain a host of terms that need further definition to clarify the original definition. He proceeds to define evaluation in six different ways, different in that each represents a distinct approach to the definitional task. (1) His classificatory definition describes evaluation as a "problem-solving strategy" employed

TM 850 521

for establishing the relative or absolute worth of various choices. (2) His comparative definition likens evaluation to research, development, management, and other problem-solving strategies, pointing out similarities and differences with each. (3) His operational definition tells how an evaluation is conducted, from identification of the impending decision through data collection and analysis to information use. (4) His compenential definition explains that evaluations include a problem, a situation involving choices, data on the worth of options, a context, a set of values, a time frame, and so on. (5) His ostensive definition gives examples of evaluations (e.g., deciding which dishwasher to buy). (6) His synonym definition includes such words as judgment and appraisal. He concludes that these six definitions, "taken together," form his concept of evaluation. He also notes that one of the difficulties encountered in sharing definitions is that, while there are at least six different ways of approaching the definitional task, "most of us fall into the habit of using only one of them" (Gephart, 1981:250-255).

Gephart's effort shows that there are various ways of approaching the definitional task. Further complicating the problem is the fact that within any one or more approaches, the content of the definition can vary. A review of a few of the variations in the content of definitions of evaluation reveals important differences in what various evaluators emphasize in their work.

(1) The classic approach of Ralph Tyler (1949) was to emphasize goals and objectives, so for him (and for the thousands of educators and researchers schooled in his approach), evaluation is the process of determining the extent to which the goals and objectives of a program are being attained.

(2) Many social scientists emphasize scientific rigor in their evaluation models, and that emphasis is reflected in their definition of the field. For these social scientists, evaluation involves primarily the application of rigorous social science methods to the study of programs (e.g., Bernstein and Freeman, 1975; Rossi, Freeman, and Wright, 1979). These evaluators emphasize the importance of experimental designs and quantitative measures.

(3) Another common emphasis in evaluation definitions is on the comparative nature of the process: Evaluation is the process of comparing the relative costs and benefits of two or more programs. The principles and definitions that undergird evaluation models emphasizing the comparative nature of the process have emerged in part as a reaction to the narrowness of evaluation when defined as measuring relative attainment of a single program's goals (see Alkin and Ellett, 1984).

(4) Still another emphasis comes from evaluators who highlight the valuation part of evaluation. From this perspective evaluation is the process of judging a program's value. This final judgment, this ultimate determination of relative merit or worth, is the sine qua non of evaluation (see Worthen and Sanders, 1973: 22-26, 120-122; Guba and Lincoln, 1981: 35-36).

(5) Some evaluation practitioners focus on the generation of data for decision making and problem solving. This perspective goes beyond making judgments or assigning relative values. The emphasis is on choices, decisions, and problem resolution. It is quite possible to decide that one thing is better than another (e.g., program X versus program Y) without taking any concrete decision with regard to program X or program Y. When evaluation is defined as a problem-solving process (Gephart, 1981) or as a

process that provides information for decision making (Thompson, 1975), some action process that goes beyond valuation is given primary emphasis in the definition.

(6) Finally, for the purposes of this discussion, there are those definitions that emphasize providing information to specific people. The broad definition I use most often takes this approach.

> The practice of evaluation involves the systematic collection of information about the activities, characteristics, and outcomes of programs, personnel, and products for use by specific people to reduce uncertainties, improve effectiveness, and make decisions with regard to what those programs, personnel or products are doing and affecting. This definition of evaluation emphasizes (1) the systematic collection of information about (2) a broad range of topics (3) for use by specific people (4) for a variety of purposes.

This definition is the basis for a "user-focused" approach to evaluation (Patton, 1981: 83-89), which places emphasis on the information needs and interests of specific people, such needs including, but not limited to, information relevant to making decisions, judgments, comparisons, or goal attainment assessments.

Now then, we have six different types of evaluation definitions (classificatory, comparative, operational, componential, ostensive, and synonym) and six different emphases in various definitions (goals, methods, comparisons, value, decisions, and information users). Nor do these cover all the possibilities. For example, in the study of how evaluations are used that formed the basis for Utilization-Focused Evaluation (Patton, 1978), I began with a collection of 170 "evaluations" on file in the Office of Health Evaluation. Fewer than half of those 170 federal health studies could be considered "evaluations" using any of the definitions just reviewed. This was because a large number of those studies were nonempirical think pieces (i.e., they included no systematic data

collection or analysis) or they focused on general social indicators without reference to any specific program. Still, they were filed (defined!) as evaluations.

Let me now make several observations based on the preceding discussion. First, no single-sentence definition will suffice to fully capture the practice of evaluation. Second, different definitions serve different purposes, one especially important function being to serve as a foundation for a particular model of or perspective on evaluation. Third, there are fundamental disagreements within the field about the essence and boundaries of evaluation. Fourth, people who propound a particular definition often have some ego investment in their special perspective, whether because they developed it, were trained according to it, or are part of a group in which that definition is esteemed; any critique of a definition, in such cases, can be taken as a personal attack, a good many people finding it difficult to separate criticism of their ideas from criticism of them personally. Fifth, people on the outside looking in (and many within the field) are often confused and uncertain about just what evaluation is. Sixth, there is no reason to expect an early end to either the disagreements or the confusion. As Samuel Butler explained the problem in "Higgledy--Piggledy,"

> Definitions are a kind of scratching and generally leave a sore place more sore than it was before.

UTILIZATION

Since there is no universally excepted definition of evaluation, <u>there can be no universally accepted definition of utilization</u>. Any given definition of utilization will necessarily be dependent on and is derived

from a prior definition of evaluation, whether that definition is implicit or explicit. As Eleanor Chelimsky has written: "The concept of usefulness...depends upon the perspective and values of the observer. This means that one person's usefulness may be another person's waste" (1983:155).

It is helpful to keep these definitional variations in mind as we review what we have learned about utilization during the last twenty years -- and I believe we have learned a great deal about the utilization of evaluation. Discussions about and research on utilization have contributed to the emergence of "utility" as one of the four central themes in the standards for evaluation developed by the Joint Committee on Standards for Education Evaluation (1981). At the professional meetings of the Evaluation Research Society and the Evaluation Network concerns about increasing the utilization of evaluation have been pervasive. We know, therefore, a great deal more than we knew a few years ago. I want to first review some of what we know and then turn to the future and identify areas about which we need to know more. To organize this discussion about what we know I shall use the six honest serving men of Kipling:

> I keep six honest serving men.
> They taught me all I knew:
> Their names are What and Why and When
> And How and Where and Who.

What Is Use

Evaluations can have conceptual or action impacts. Conceptual impacts are those which affect thinking about a program. Such uses may lead to conceptualizing implementation or outcomes in new ways, understanding dynamics of the program more thoroughly, or shifts in program priorities. Action impacts are those which lead to observable changes in the actual operations of a program. These are most notable and dramatic when they

involve changes in levels or types of funding, or changes in program delivery. Evaluations can also affect decisions. Evaluation may lead to a decision to continue or stop a program, or to do any of the large variety of things over which decisionmakers have control. A decision to do absolutely nothing new or different can be a major evaluation impact but will not lead to any observable action or change as a result of the evaluation.

The research on utilization is typically biased towards action impacts. The early litanies about the lack of utilization of evaluation were based on narrow definitions of utilization limited to immediate action. It is clear, though, from discussions with people who actually use evaluations (Patton, 1978) that reinforcing or challenging ways of thinking are important impacts for decisionmakers attempting to reduce their uncertainties about programs.

The relative importance, then, of evaluation utilization can only be judged perceptually by the value attached to utilization by those who use the evaluation. There can be no absolute standard which values action over thinking, changes in a program over keeping things the same, or decisions to do something over decisions to wait. There simply can be no hierarchy of impacts because the hierarchy is necessarily situational and depends on the values and the needs of the people for whom the evaluation is conducted.

What Is Used

Early research on utilization focused on the outcomes of evaluation, i.e., the data, the recommendations, and the evaluation report. When the question was asked, "Was the evaluation useful?" the implicit assumption was that one was talking about the findings and recommendations of the

evaluation.

As our understanding of the utilization process has increased, however, we have come to understand that evaluation processes can have significant impact quite apart from the outcomes of the evaluation. Indeed, evaluation processes can be used even if there are no outcomes of an evaluation, for example, if data collection falls apart and no report is ever written. Evaluation processes can be useful in helping program staff clarify what they are doing, establish priorities, focus resources and activities on specific outcomes, and identify areas of weakness even before data are collected. Evaluation processes are useful because they stimulate staff to think rigorously about their program in ways which might not happen without the forced stimulus of coming to grips with the demands of the evaluation.

In addition to the use of evaluation findings and evaluation processes, there is the use of the evaluator. Quite apart from facilitating evaluation processes and producing outcomes, the evaluator can be useful in program development and decision-making as a professional consultant who is sensitive to and insightful about the program. This utilization of the evaluator can be viewed quite apart from and go well beyond the more narrow confines of data collection and goals clarification. It is not unusual for evaluators to be asked for their perceptions, their impressions, their managerial assistance, and their general advice above and beyond the narrow focus of data collection. Some evaluators are uncomfortable with this larger role and refuse to take it on, but many evaluators consciously or unknowingly become management consultants and important advisors to programs.

Who Uses Evaluation

There are multiple and varied interests around any evaluation. Teachers, administrators, parents, public officials, funders, students, and community leaders all have an interest in educational evaluations, but the degree and nature of their interests will vary. We have learned that these different constituencies use evaluations in different ways. Program staff are most likely to benefit from utilization of the evaluation process. Funders and the community people are most likely to use published data and written findings. Administrators are most likely to use the evaluator as a consultant. The kind of impact also varies. An evaluation is likely to be used to affect the thinking and conceptualization of people more distant from the day-to-day operations of the program, to affect actions taken by those actually involved in the day-to-day delivery of the program, and to affect the decisions taken by those with overall responsibility for the program, i.e., funders and administrators.

We have also learned that the extent to which these various constituencies are well served will vary from evaluation to evaluation. No single evaluation is likely to be able to serve all constituencies equally well. Either implicitly or explicitly the evaluation design includes bias toward the information and process needs of some constituencies more than others. We have also learned that we should include among the list of possible constituencies who benefit from the evaluation the evaluator himself or herself. An evaluation may serve the needs and interests of the evaluator more than those of any other constituency. Whether such use is acceptable, justifiable or important depends on where one is sitting and what values are brought to bear on the question of utilization. Nor am I talking here simply about lining the pockets of evaluators or providing

academics with publishable papers. I recently talked with a state administrator of welfare programs who had created an internal evaluation unit at the state level. During the first year of operations, the evaluation unit had conducted several evaluations all of which the state administrator judged to be useful "because they helped the evaluators learn how to conduct evaluations at the state level." None of these evaluations had yielded particularly useful information for the state administrator, nor were the subjects of the evaluation important for the state. The purpose of the evaluations conducted during the first year were entirely aimed at making the unit operational so that it could be useful in generating important data for decision-making in the future. This is a case of utilization of evaluation processes for the benefit of evaluators, at least in the short run.

When Is Evaluation Used

The qualifying phrase at the end of the last sentence points out the problem of determining a time horizon for the utilization of evaluations. The early literature on utilization of evaluations focused on immediate action impacts. Subsequent research found that evaluation utilization was more likely to be incremental than immediate. This means that, in many cases, evaluation processes make a difference time and that evaluation outcomes (findings) are discussed and used over a period of time. This incremental nature of evaluation utilization flows in part from the incremental nature of most decision making. There are not a great many clear, specific and immediate decisions taken in public organizations. Rather, decision making tends to be a process of moving in a particular direction that is not always explicit and does not always come from decisive moments of action. There remains, I believe, a bias in the

research literature on utilization in the direction of prefering immediate, concrete and short-term impacts to more diffuse and longer term impacts. This is partly in response to the measurement problem, i.e., that more immediate impacts are easier to get at and are more visable. However, incremental impacts over a longer period of time may be more important in many cases.

How Is Evaluation Used?

There are many dimensions one might consider here. I want to focus on two, more by way of example than because they are definitive, although they have been particularly importart in the evaluation utilization literature. Evaluation utilization can be planned or unplanned, and can be formal or informal. Planned utilization occurs when the intended use of the evaluation is identified at the beginning and then subsequent utilization follows and is judged by planned or intended use. Unplanned utilization occurs when, in the typical case, the evaluation is designed without particular attention to questions of utilization and questions of use are left until the data are collected and analyzed. Eleanor Chelimsky argues that the most important kind of accountability in evaluation is utilization that comes from "designed tracking and follow-up of a predetermined use to predetermined user" (1983:160). Chelimsky calls this a "closed-looped feedback process" where "the policy maker wants information, asks for it, and is interested in and informed by the response" (1983:160). From this perspective, the most important question in researching the utilization of evaluation is whether the evaluation had its intended use. This solves the problem of defining utilization, addresses the question of who the evaluation is for, and builds in a time frame since the predetermined use would necessarily have a time frame.

The problem here is the same problem that emerges in evaluation itself when attention is directed only to the stated and explicit goals of the project. The debate that led to Scriven's proposal for goal free evaluation included concerns about attention to unanticipated consequences, side effects, and unstated goals as important outcomes of programs. Attention only to explicitly stated goals would miss these other impacts. Likewise, in looking at utilization, limiting attention to the explicitly stated expectations for utilization will miss longer term, unintended, and unplanned uses, any of which may be quite important.

Another aspect of how evaluations are used that has become important is whether uses formal or informal. The early research on utilization focused on formal uses, that is public, observable, and explicit uses of published findings. We have since learned that informal uses are often more important. This is the transfer of findings by word of mouth, in unplanned discussion groups, and in one-to-one interactions between the evaluator and program staff, administrators, and/or funders. Such informal interactions often go well beyond official evaluation findings, and it is in the informal process of utilization that the evaluator himself or herself is likely to be used as much as or more than either the formal evaluation process or findings.

Where Is Evaluation Used?

The problem of where evaluation is used has emerged most directly in efforts at satisfying the different needs of evaluation users at the local, state, and national levels in education. Framed in this way the question of where is closely related to the question of who. But the question of where the evaluation is used is a larger dimension in that evaluation designs and potential uses at the national level are quite different from

those aimed at local utilization. In a perfect world, the kind typically demanded by political rhetoric, a single evaluation would be useful at all levels from the local school up through the national government. In reality, the information needs of these different units are dramatically different. Indeed, the most common problem I encounter in evaluation consulting is dealing with the conflicting information needs of people at different levels of government. The state system imposes data collection requirements on local units that they preceive to be useless while data collected entirely by local initiative seldom meets the needs of either state or federal governments. Local units tend to prefer highly idiosyncratic and situationally specific data. Larger units tend to prefer standardized data which makes aggregation and comparisons easier. All of the dimensions of utilization vary according to where the evaluation is used. One of the greatest challenges for evaluations that are part of management information systems is responding to utilization needs at these different levels.

Why Is Evaluation Used?

The "why" of evaluation use has focused most often on the distinction between formative and summative evaluations. Indeed, the classic formative-summative distinction was intended to define different kinds of evaluation use, i.e., evaluations aimed at program development and improvement versus evaluations aimed at major go/no-go decisions and/or major funding decisions. In practice, however, the "why" question is considerably more complex than this. The reasons evaluations are used, or not used, run the gamut of human motivations and schemes. There are highly political reasons why evaluations are used or not used. There are personality dimensions to this problem. There are personal value reasons,

and matters of personal integrity and motivation. There are reasons having to do with human factors, context factors, and characteristics of the evaluation. Indeed, the question of why evaluations are used leads directly to the research literature which reports on the factors which affect evaluation use, which explain utilization, and which describe varying conditions under which utilization takes different forms. Indeed, most of the research on utilization has focused on identifying the factors that contribute to use rather than on variations in utilization itself, this later point having been the focus of my discussion thus far.

James Burry (1984) has done a thorough review of the evaluation utilization literature aimed at a synthesis of factors which appear to have a bearing on the degree to which evaluation information may be used. He organizes the various factors in three major catagories: human factors, context factors, and evaluation factors.

> Human factors reflect evaluator and user characteristics with a strong influence on use. Included here are such factors as people's attitudes toward and interest in the program and its evaluation, their backgrounds and organizational positions, and their professional experience levels.
>
> Context factors consist of the requirements and fiscal restraints facing the evaluation, and relationships between the program being evaluated and other segments of its broader organization and the surrounding community.
>
> Evaluation factors refer to the actual conduct of the evaluation, the procedures used in the conduct of the evaluation, and the quality of the information it provides (Burry, 1984).

Since I presume that the participants in this meeting will have access to the Burry review, I shall not discuss it in greater detail at this time. That review, in conjunction with and as a part of the framework developed by Alkin et al (1979), presents a comprehensive look at the factors affecting evaluation. (In accepting the invitation to do this paper I had orginally expected to do what I subsequently found James Burry had done.

His work is so well done I saw no reason to replicate his efforts.)

The primary weakness of the synthesis and of the framework developed by Alkin et al (1979) is that the factors are undifferentiated in terms of importance. The synthesis represents a checklist of factors which can influence evaluation, and the literature which is synthesized suggests the conditions under which certain factors will emerge as important, but no overall hierarchy is suggested by the synthesis, i.e., a hierarchy which places more importance on certain factors as necessary and/or sufficient conditions for evaluation utilization. In the next section I want to take on this problem of differentiating the relative importance of various factors which explain utilization.

THE PERSONAL FACTOR

I want to suggest that the personal factor is the most important explanatory variable in evaluation utilization. I make this assertion quite deliberately in order to be provocative. The personal factor emerged as the most important variable in the initial research that led to Utilization-Focused Evaluation (Patton, 1978). A great deal of subsequent research has validated the importance of this factor, and I know of no research which would indicate that it is not the key variable in utilization.

The personal factor has to do with the interests and commitments of the key people involved in the evaluation. Where the key people are interested in, committed to, and involved in the evaluation for the purpose of making sure that it is useful, then the evaluation is likely to be used. Where those interests, commitments and involvement are not present, evaluation is considerably less likely to be used.

The personal factor is general in conceptualization. It includes several, but not all, of the factors listed in the Burry synthesis as "human factors." However, I prefer the term "personal factor" to "human factors" because the phrase, "the personal factor," is meant to explicitly communicate that the personal characteristics of individual people is what makes the difference. This is in strong contrast to structural, organizational, and methodological explanations.

Identification of the personal factor as the key explanatory variable also provides a bridge from description to prescription. Given that the standards for educational evaluations (1981) have included a clear mandate for evaluator accountability which holds forth the ideal that evaluations first and foremost should be useful, it seems to me appropriate to use our knowledge of factors affecting the utilization of evaluation to make prescriptive statements that will guide evaluators in their efforts. Such prescription goes beyond saying that one ought to take a certain list of factors into account when designing the evaluation. A prescriptive statement that is research-based would tell evaluators <u>how</u> to take those factors into account. The personal factor provides such a prescription by saying one takes those factors into account in terms of the values, interests, and understandings of the people who are to use the evaluation results.

For example, I began this paper by reviewing the problem of different definitions of evaluation. A neutral checklist approach to evaluation practice would advise the evaluator to understand that there are different definitions of evaluation and to be sure to define what kind of evaluation is being undertaken in a particular effort. However, a utilization-focused approach built on the importance of the personal factor provides more

direct advice about how to proceed. A utilization-focused approach begins by finding out the perceptions and definitions of the people with whom one is working. Before the evaluator unilaterally defines evaluation, the evaluator should work to discover the perceptions, confusions, expectations, and beliefs about evaluation of those people who will be the primary users of the evaluation. It is then possible to build on that knowledge to develop shared understandings about evaluation options and potential processes. It is often appropriate simpdy to ask the people with whom one is working to associate freely in a stream-of-consciousness fashion with the word earticipants to define evaluation. The question, "Who can give me a definition of evaluation?" clearly implies a single right answer, and the wary participant will suspect that the evaluation trainer or facilitator will eventually pronounce the correct definition, but only after making several participants look stupid. Definitions are thus perceived as academic playthings to be used in a game at which the researcher is sure to win, so why participate? I'm not looking for skill at constructing or repeating definitions. I'm looking for perceptions and synonyms that will provide clues to tacit definitions held by people in the situation in which I'm looking. With these perceptions made explicit it is then possible to consider other alternatives and end by defining evaluation in a way that is relevant to the people who are going to use the evaluation.

In bridging the gap between description and prescription, the gap between a list of possible factors and the more direct identification of how those factors come into play, it is possible to use our knowledge of evaluation utilization to state a set of premises or prescriptions about how to increase utilization. Those premises constitute what I have called

"utilization-focused evaluation." I am presenting them here as a way of testing the extent to which we agree that there is a research base for these premises.

UTILIZATION FOCUSED EVALUATION PREMISES

1. Basic Premises of Utilization-Focused Evaluation

The first premise is that concern for utilization should be the driving force in an evaluation. At every point where a decision about the evaluation is being made, whether the decision concerns the focus of study, design, methods, measurement, analysis, or reporting, the evaluator asks: "How would that affect the utilization of this evaluation?"

The second premise is that concern for utilization is on-going and continuous from the very beginning of the evaluation. Utilization isn't something one becomes interested in at the end of an evaluation. By the end of the evaluation, the potential for utilization has been largely determined. From the moment decisionmakers and evaluators begin conceptualizing the evaluation, decisions are being made which will affect utilization in major ways.

The third premise is that evaluations should be user-oriented. This means that the evaluation is aimed at the interests and information needs of specific, identifiable people, not vague, passive audiences. Therefore, the first step in utilization-focused evaluation is identification or organization of specific decision makers and information users. The evaluator must determine who the potential users are, and aim the evaluation at those users.

A fourth premise is that, once identified, these interested decision makers and information users should be personally and actively involved in

making decisions about the evaluation. Working actively with people who have a stake in the outcomes of an evaluation (the "stakeholders") is aimed at increasing the potential for utilization by building a genuine commitment to and understanding of the evaluation over the course of the evaluation process. Such an approach recognizes the importance of the "personal factor" (Patton, 1978) in evaluation utilization. People who are personally interested and involved in an evaluation are more likely to use evaluation findings. The best way to be sure that an evaluation is targeted at the personal concerns of stakeholders is to involve them actively at every stage of the evaluation.

A fifth premise is that there are multiple and varied interests around any evaluation. Teachers, administrators, parents, public officials, students, and community leaders all have an interest in evaluation, but the degree and nature of their interests will vary. The process of identifying and organizing stakeholders to participate in an evaluation process should be done in a way that is sensitive to and respectful of these varied and multiple interests. At the same time, it must be recognized that resource, time, and staff limitations will make it impossible for any single evaluation to answer all possible questions, or to give full attention to all possible issues. Identified decision makers and information users, representing various constituencies, should come together at the beginning of the evaluation to decide which issues and questions will be given priority in the evaluation in order to maximize the utility of the evaluation. The process of focusing the content of the evaluation should not be done by evaluators acting alone, or in isolation from users.

A sixth premise is that careful selection of stakeholders for active participation in the evaluation process will permit high quality

participation, and high quality participation is the goal, not high quantity participation. The quantity of group interaction time is often inversely related to the quality of the process. Thus, evaluators conducting utilization-focused evaluations must be skilled group facilitators and have a large repertoire of techniques available for working actively with stakeholders in the evaluation (Patton, 1981). High quality involvement of stakeholders will result in higher quality evaluations. Many evaluators assume that methodological rigor will inevitably be sacrificed if non-scientists collaborate in making methods decisions. This need not be the case. Decision makers want data that are useful and accurate (Weiss and Bucuvalas, 1980). Skilled evaluators can help non-scientists understand methodological issues so that they can judge for themselves the trade-offs involved in choosing among the strengths and weaknesses of design options and methods alternatives. Such involvement in collaborative deliberations on methodological issues can significantly increase stakeholders' understanding of the evaluation, while giving evaluators a better understanding of stakeholder priorities and situational constraints on the feasibility of alternative approaches. These shared decisions can thus enhance both utilization potential and methodological rigor.

A seventh premise is that evaluators committed to enhancing utilization have a responsibility to train decision makers and information users in evaluation processes and the uses of information. By training stakeholders in evaluation methods and processes, the evaluator is looking to both short-term and long-term utilization. Making decision makers more sophisticated about evaluation can contribute to greater use of evaluation data and evaluation processes over time.

An eighth premise is that there are a variety of ways in which evaluation processes and findings are used, a point noted earlier. Evaluations can directly influence major, specific decisions. Evaluations can be used to make minor adjustments in programs. Decision makers can, and do, use evaluations to reduce uncertainty, enlarge their options, increase control over program activities, and increase their sophistication about program processes. Sometimes evaluations have more of a conceptual impact, i.e., they influence how stakeholders think about a program, rather than an instrumental impact, i.e., evaluation utilization manifested in concrete actions and explicit decisions. A broad view of utilization reveals multiple layers of impact over varying amounts of time. All of these kinds of utilization are important and legitimate form a utilization-focused evaluation perspective. This view of utilization also broadens the notion of evaluation impact to include use of the entire evaluation process as a stakeholder learning experience, not just use of the findings in the final report. The relative valuae of these different kinds of utilization can only be judged in the context of a specific evaluation. There is no universal hierarchy where some kinds of use are always more valuable.

A ninth premise is that attention to utilization involves financial and staff time costs that are far from trivial. The benefits of these costs are manifested in greater utilization. These cost should be made explicit in evaluation proposals and budgets so that utilization efforts are not neglected for lack of resources.

A tenth premise is that a variety of factors affect utilization. These factors include community variables, organizational characteristics, the nature of the evaluation, evaluator credibility, political considerations,

and resource constraints (Alkin et al, 1979) In conducting a utilization-focused evaluation, the evaluator attempts to be sensitive to and aware of how these various factors affect the potential for utilization. An analysis of the factors that may affect the usefulness of an evaluation should be undertaken jointly with stakeholders early in the evaluation process. These factors, and their actual effects on utilization, are then monitored throughout the utilization-focused evaluation process.

TESTING THESE PREMISES IN PRACTICE

The integrated nature of these premises with the underlying importance of the personal factor suggests to me the primary direction for future research on utilization. That direction is to test these premises in an integrated way in real evaluations. By an integrated way I mean that it is important to look at utilization in terms of the related factors and processes that occur from beginning to end from a fairly holistic perspective. Studies that focus on one or two single, isolated factors are of little use in extending our knowledge of real utilization processes. Laboratory experiments aimed at testing out what happens if people have more or less information, or some other single factor taken in isolation from real world settings, is also of little use. The premises stated above mean that utilization research must necessarily be holistic and must look at entire evaluation processes from beginning to end. There is no other way to test out these premises because they are not subject to simple manipulation of operational variables.

By way of illustrating how such tests can proceed, I would like to describe a recent evaluation project which employed a utilization-focused

evaluation perspective and permitted what has been in my experience the most comprehensive test of utilization-focused evaluation in a real world project. That project is the Caribbean Agricultural Extension Project for which I act as Project Director. Thus, my account of the external evaluation of that project and its use is necessarily subjective. However, even as this paper is being written the chair of this external evaluation is also independently reviewing the utilization process.

A CASE EXAMPLE OF UTILIZATION

The Caribbean Agricultural Extension Project is a U.S.AID-funded project aimed at improving national agricultural extension services in eight Caribbean countries. With staff from the University of Minnesota and the University of the West Indies, the project has involved organizational development work with key officials in the eight countries, providing in-service training for extension staff, and providing equipment including vehicles, office equipment, and agricultural equipment.

The project was designed based on 1-1/2 years of needs assessment and planning. The assessment and planning included establishing an advisory committee in each country as well as a regional advisory committee made up of representatives from all eight participating countries and other organizations involved in agricultural development in the Caribbean.

The contract for implementing the project with all key participants was signed in January, 1983. In April, 1983, a meeting of the Regional Advisory Committee was held with a team of external evaluators. The external evaluators were chosen to represent the major constituencies of the project, these being (1) U.S.AID, (2) the University of the West Indies, and (3) the Midwest Universities Consortium for International

Activities (MUCIA) for which the University of Minnesota was the primary representative. Each of these three prime constituencies named one of the evaluators. The fourth evaluator was chosen for his stature in the field of evaluation, because of his commitment to user-oriented evaluations, and because he was neutral from the point of view of the other three constituencies. He was made chair of the evaluation team so as to represent neutrality in collaboration with the three evaluators who had been named by specific constituencies as able to represent their points of view.

Prior to designing the evaluation, the evaluators met with representatives each of these constituencies separately, including the funding source U.S.AID. At the April meeting of the regional advisory group, the evaluators focused three days of discussion on the criteria which could be used to determine if the project had been successful. These criteria constituted a set of questions and primary outcomes, but were not quantitative indicators. Based on those discussions the evaluators reviewed design possibilities with the fifty participants in that regional advisory meeting. The details of the design were then worked out with specific representatives of the project staff and U.S.AID.

The evaluation design included several different foci. The project staff organized all of their required reporting around the evaluation design. The work plan for project staff was also developed based on the evaluation elements and staff meetings routinely reviewed the elements of the evaluation as a way of directing implementation and focusing on those outcomes which were primary from the point of view of the project and the evaluation. Members of the evaluation team were sent monthly and quarterly reports based on the elements of the evaluation. For example,

the first element in the evaluation design was that in each country a national agricultural extension planning committee be operating and involved in providing direction to the extension service in its country. All staff meetings began by reviewing the progress of national planning committees and all monthly and quarterly reports include information on the activities and progress of national committees. In addition, the minutes of the national planning committees were provided to the evaluators. In the actual data collection phase the evaluators conducted interviews to gather firsthand information about the operations and activities of the the national planning committees. The point here is that program implementation and evaluation were synchronized from the beginning of each. More importantly, the evaluation process had a major impact on improving program implementation from the very beginning by focusing program implementation. The evaluation provided a framework for program planning and reporting that provided focus to staff activities. This focus became more important as the project moved forward and staff encountered many opportunities to be diverted from those primary foci. However, having organized the project work plan, staff meetings, and reporting around the key evaluation elements, the evaluation contributed substantially to keeping staff efforts from being diverted into other areas or activities which would have taken away from the primary purpose of the project. This is an example of utilization of evaluation processes for program improvement.

Data collection and reporting were carefully timed to provide critical information for refunding decisions. Working backwards from the project completion date, a time schedule for data collection and reporting was developed which would make sure that the information was available when the

decision about refunding and future project activities was to be made. This was a major break with U.S.AID tradition. Indeed, the evaluators and project staff had some difficulty helping U.S.AID personnel understand why the evaluation was taking place so early in the life of the project, early from their perspective. Traditionally, U.S.AID evaluations occur after a project is completed to provide a mandatory report on project impacts. That means that the traditional U.S.AID evaluation is presented six months to a year after a project is terminated or a new funding decision has been made. There is no possibility of the evaluation playing a role in that decision making. It was unprecedented for U.S.AID to get an evaluation report, at least one that was more than cursory, at the time of a funding decision. In this case operational project funding would end in September, 1985. Given the lengthy funding process of U.S.AID a decision for additional funding and activities would have to be made by December, 1984, to do the paperwork to keep the project alive. Thus, a meeting of the Regional Advisory Group was scheduled for November, 1984, to focus on the evaluation findings. This meant that the report would have to be ready by that time so the data collection would have to take place in the Summer of 1984, only 1-1/2 years into project implementation and only a year after the initial design, fully a year ahead of the operational project completion date. Clearly, such an evaluation could not be definitive about project impacts since data collection would take place well before project completion, but a definitive data collection effort would not be available at the time the decision was to be taken. Data collection did occur in June of 1984 and the evaluation report was ready for the Regional Advisory meeting in November, 1984. Prior to that critical November Regional Advisory Committee Meeting, the evaluators met separately with project

staff to provide informal feedback about evaluation findings and with U.S.AID to provide informal feedback and discussion of potential future funding. In both cases those informal meetings were critical.

The first informal meeting occurred immediately after data collection in June, 1984. The evaluators, who had been gathering data in different locations, met together to review their findings and divide the writing tasks. Following that session together, the evaluators met with the project director to review major findings. Those findings included a confirmation of the overall successes of the project, the high degree of support for project activities among the participating countries, and identification of areas of weakness. The areas of weakness included insights which had escaped the attention of project staff. The staff immediately began to correct those weaknesses, two of which required assistance from outside and one of which brought a new focus to implementation activities. A month later one representative of the evaluation team met with the project staff in their full staff meeting and reviewed the evaluation findings. It was at that staff meeting that activities were reoriented to direct attention to identified weaknesses.

Following the staff meeting, the evaluator who had been selected by U.S.AID met with U.S.AID officials to informally report initial findings. At that meeting the question of future funding arose. The director of the funding agency, U.S.AID, had been present in the initial meeting with the evaluators where important questions were identified. He now put those questions to the evaluator again with special reference to future activities. The evaluator was able to directly address this question with high credibility and with concrete data. It has subsequently been reported to me independently by several U.S.AID staff that this informal feedback

was critical because the director of the funding agency was not predisposed to continue funding for the project. The evaluation report made it clear that the project was effective, was having an impact, but that further funding and activities would be necessary and justifiable to institutionalize short term successes and guarantee long term success and long term effectiveness. With the November, 1984, Regional Advisory meeting already scheduled, and with the informal evaluation results having been reported to the funding agency, I wrote to U.S.AID asking for them to take a position on their openess to future funding, I asked for a response prior to the November, 1984, meeting of the Regional Advisory group since the delegates to the Regional Advisory Group would need to know AID's position as a context for their discussions of the evaluation. Prior to that meeting U.S.AID indicated that they had reviewed the evaluation and were inclined to continue funding activities. They therefore invited project staff and the Regional Advisory group to submit a continuing proposal.

The published evaluation report was completed in time for the November Regional Advisory meeting. At that meeting the evaluators reviewed overall findings and different ways in which the report could be used for local purposes as well as regional purposes. Delegates to the meeting reviewed the executive summary and commented on its accuracy. They then adopted a resolution accepting the evaluation report as generally accurate, fair, thorough, and balanced. They suggested that project staff use the evaluation findings as a basis for future activities and a new proposal to U.S.AID. They discussed major new directions suggested by the evaluation findings. They brought to bare on those discussions other information and their own experiences, and subsequently adopted resolutions identifying the

major components that should be included in continuing activities.

In the interim between the data collection, informal feedback and the formal November review of the evaluation, project staff had made major progress in overcoming the weaknesses identified in the preliminary feedback. In addition, through the grapevine, the fact that the evaluation report would show substantial progress and major successes was communicated throughout the region. Project staff and U.S.AID had the opportunity to comment on draft copies of the report before it was published to guarantee accuracy and so as to know details of what the report would say prior to its publication.

Thus, the evaluation had a major impact on project implementation. It had a major corrective effect in reorienting the project a year and a half into implementation so as to correct weaknesses that had emerged during that time and to more directly focus on some areas that were being neglected. Finally, the evaluation had a major impact on the decision to continue project funding.

The evaluation direct costs were approximately $100,000 out of a total project budget of $5.4 million dollars. This is under two percent of the project budget.

While the details of this evaluation are skimpy to preserve space, all of the premises of utilization-focused evaluation were followed in this evaluation and the result was a high level of use. Of course, this design does not permit one to make causative statements about the relationship between what was done and what subsequently occurred. However, there is no question among the nine project staff members nor the U.S.AID officials that both the evaluation processes and outcomes made important differences. Likewise, the resolution adopted by the Regional Advisory group made it

clear that they had learned from the evaluation both about the project and about how evaluations ought to be conducted.

Issues Needing Clarification

This discussion leads me to five issues that continue to create confusion in the field. At the recent ERS-ENet meetings in San Francisco I found evidence that these issues are alive and well. In some cases it seems to me we can put them to rest. In other cases new work is needed. These issues are:

1. What is the relationship between quantity and quality of interaction between evaluators and decision makers?

2. How does heavy involvement of stakeholders in an evaluation affect methodological quality?

3. What, if any, is the hierarchy of desired impacts from an evaluation i.e., is immediate action a "greater impact" than long term effects on program thinking and conceptualization?

4. Is there a hierary in terms of the parts of an evaluation that ought to be used, i.e., is use of findings more important than use of the process, and are both more important than use of the evaluator?

5. Is predetermined use for predetermined users more important than unintended use in unintended ways?

Permit me to elaborate on the background which gives rise to these questions. The importance of the personal factor and its manifestation in "stakeholder-based evaluations" has led to confusion about the nature of evaluator involvement with key decision makers and information users. This is sometimes called "the stakeholder assumption."

The "stakeholder assumption" is the idea that key people who have a stake in an evaluation should be actively and meaningfully involved in shaping that evaluation so as to focus the evaluation on meaningful and appropriate issues, thereby increasing the likelihood of utilization. A consultative evaluation approach is based on the stakeholder assumption.

In recent years, as evaluators have become increasingly concerned about utilization, the stakeholder assumption has received widespread attention. Doubts have been raised about the validity of the assumption. Nick Smith, for example, president of the Evaluation Network during 1980, wrote in his column in the Evaluation Network Newsletter that the assumption was being accepted without sufficient empirical evidence to support the supposed relationship between stakeholder involvement and utilization of findings.

> Although this (the stakeholder assumption) appears to be a widely held belief, no one has bothered to test it empirically. From a recent 16-state study of local district school accreditation evaluations, I have found that data from school board members and administrators with first-hand experience in such evaluations do not agree with this assumption. These individuals do not want to be personally more involved in such studies, nor do they believe that their involvement will make the evaluation results more useful to them. In fact, for these school board members and administrators, the correlation between their judgments of a past evaluation's utility to them was only 0.3, while there was a 0.7 correlation between their judgments of the evaluation's quality and its utility. Hardly strong support for the considerable effort now being expended at the local, state, and federal level to increase the involvement of various groups in evaluation (Smith, 1980: 39).

Smith's doubts about the validity of the stakeholder assumption provide an opportunity to clairfy my own interpretations about what the stakeholder assumption means in practice. His critique includes some common misconceptions about the collaborative approach to utilization-focused evaluation which give rise to the issues outlined above.

First, there is the question of the nature of the relationship between stakeholder involvement and utilization. Smith states the relationship as a "necessary" condition. In the sentence preceding the excerpt quoted above, he said he was addressing "the currently popular assumption that increased involvement of clients and decision makers in evaluation

activities will necessarily result in increased utilization of evaluation findings" (Smith, 1980: 39, emphasis added). From my point of view, the stakeholder assumption is somewhat overstated by Smith. I have never suggested, or heard others suggest, that increased stakeholder involvement in an evaluation will necessarily result in increased utilization. Nothing one can do, as near as I can tell, will guarantee utilization.

Second, there is a hint of a tradeoff in Smith's skepticism implying that one must choose between stakeholder involvement and high quality data. Many evaluators assume that methodological rigor will inevitably be sacrificed if nonscientists collaborate in making methods decisions. This need not be the case. The ideal expressed in the new standards includes both utility and accuracy. Other research confirms Smith's findings that decision makers are concerned about "quality" of data, but quality includes both "truth tests" (accuracy) and "utility tests," the latter being a concern for relevance and applicability (Weiss and Bucuvalas, 1980).

Third, Smith's point is directed entirely at the quantity of stakeholder involvement in an evaluation. The variable "level of involvement" is somewhat ambiguous, but the implication is that it refers to amount of involvement in terms of time. In contrast, the emphasis in utilization-focused evaluation is on careful selection of the people with whom one works and the quality of the evaluation group process. The quantity of interaction time is often inversely related to quality.

Fourth, while I expect Smith is correct in saying that there is a dearth of empirical evidence that "increased involvement" (if he means greater quantity) leads to greater utilization, there is substantial evidence that high-quality involvement of the right people contributes substantially to utilization. The massive diffusion of innovation

literature (e.g., Havelock, 1971, 1973; Rogers and Shoemaker, 1971) is replete with relevant empirical evidence. The formal organizations, participatory management, and small group literatures in psychology and sociology provide substantial data relevant to this point (e.g., Hage and Aiken, 1970; Bennis, 1966; Azumi and Hage, 1972; Bennis et al., 1976; Argyris, 1972, 1974, 1976). These literatures document with empirical evidence the proposition that people are more likely to accept and use information, and make changes based on information, when they are personally involved in and have a personal stake in the decision making processes aimed at bringing about change. Most directly, there is a growing evaluation and policy analysis literature--an empirical literature--that supports the proposition that utilization of evaluation is enhanced by high-quality stakeholder involvement in and commitment to the evaluation process (e.g., Fairweather et al., 1974; Weiss, 1977; Patton, 1978; Alkin et al., 1979; Braskamp and Brown, 1980; Stevens and Tornatsky, 1980; Lynn, 1980; Dickey, 1981).

Fifth, evaluators should not expect much initial enthusiasm among stakeholders for the idea of participating actively in a research process. Past experiences are not likely to have been very positive. Most stakeholders are quite happy to leave evaluation to evaluators. They're also quite happy to ignore the resultant evaluation findings. Like a child who wants to avoid bad-tasting medicine (medicine, by definition, being bad tasting), stakeholders would typically prefer to avoid being subjected to distasteful doses of evaluation (evaluation, by definition, being distasteful), even if they believe it's good for them. The evaluation practitioner, like the medical practitioner, must often cajole and otherwise persuade stakeholders to do what ought to be done. Getting

cooperation and participation has to be worked at. Initial resistance is no reason to fall back on traditional patterns of operating alone, at least not if the evaluator is really committed to utilization. In my experience, if stakeholders won't get involved at the beginning of an evaluation, they probably won't pay it much heed at the end.

If utilization is viewed as use of the evaluation process and not just as final findings, then stakeholders must be involved in the entire process for the process to have the most impact. Much of the impact of evaluations on stakeholders comes through personal engagement in the difficult processes of goals clarification, issues identification, operationalizing outcomes, matching research design to program design, determining sampling strategies, organizing data collection, interpreting results, and drawing conclusions. These processes take stakeholders through a gradual awakening to program complexities and realities, an awakening that contains understandings and insights that will find their way into program developments over time, only some of which will be manifested in concrete decisions. Utilization begins as soon as stakeholders become actively involved in evaluation because that involvement, properly facilitated, forces them to think about program priorities and realities. The stakeholder assumption, then, includes the expectation that stakeholders need to expend time and effort to figure out what is worth doing in an evaluation; they need help in focusing on worthwhile questions; and they need to experience the full evaluation process if that process, which is really a learning process, is to realize its potential, multi-layered effects.

To summarize, there are several key issues that emerge from discussions of the stakeholder assumption and research on the personal

factor:

1. What is the relationship between quantity and quality of interaction between evaluators and decision makers?

2. How does heavy involvement of stakeholders in an evaluation affect methodological quality?

3. What, if any, is the hierarchy of desired impacts from an evaluation i.e., is immediate action a "greater impact" than long term effects on program thinking and conceptualization?

4. Is there a hierarchy in terms of the parts of an evaluation that ought to be used, i.e., is use of findings more important than use of the process, and are both more important than use of the evaluator?

5. Is predetermined use for predetermined users more important than unintended use in unintended ways?

These issues emerge from the implicit hierarchies and values that seem to me to be present in the current research on evaluation. These are not simply research issues. Rather, they have to do with how the research on utilization ought to be conceptualized for future work.

Before closing this paper, I would like to raise one additional issue for discussion. That issue is the misutilization of evaluations.

MISUTILIZATION

If it is not already clear to the reader, let me make it absolutely clear: I believe that we already know enough about how to increase the utilization of evaluations that the immediate task is acting on what we know and evaluating those actions rather than doing further isolated research on utilization. In so acting, and in order to be accountable, evaluators ought to document their experiences in using what we know. In presenting my recent experiences with the Caribbean Agricultural Extension Project as a test of utilization-focused evaluation premises, I have indicated in very brief fashion how such documentation might take place.

For my part, however, I don't look for any major new breakthroughs in research on the utilization of evaluation. I'm satisfied that if we actually begin to use what we already know, we could make a significant difference in evaluation practice.

I have become, however, increasingly concerned about problems of misutilization. The most common criticism I hear of utilization-focused evaluation is that it co-opts evaluators. This cooptation reduces evaluation credibility, neutrality, and significance.

In our concern with and focus on ways of increasing the utilization of evaluation, I agree with those who worry that we have neglected misutilization. As I do workshops and travel around the country talking with people about evaluation, I hear increasingly about cases of abuse and misuse. As I've thought about this, I'd like to share some preliminary observations by way of generating additional discussion on this important issue.

1. Misutilization is not at the opposite end of a continuum from utilization. There are really two dimensions here. One dimension is a continuum from nonutilization to utilization. A second continuum is non-misutilization to misutilization. Studying misutilization is quite different from studying utilization.

2. Having conceptualized two separate dimensions, it is possible to explore the relationship between them. Therefore, permit me the following proposition: <u>as utilization increases, misutilization will also increase</u>. It seems to me that when people ignore evaluations, they ignore their potential uses as well as abuses. As we successfully focus greater attention on evaluation data, and as we increase actual use we can also expect there to be a corresponding increase in abuse, often within the same

evaluation experience.

3. Misuse can be either intentional or unintentional. Unintentional misutilization can be corrected through the processes aimed at increasing appropriate and proper utilization. Intentional misutilization is an entirely different matter to which, it seems to me, we have paid very little attention except to say it shouldn't happen. In terms of incidence and prevalence, I have no clear notion of whether unintentional or intentional misuse is more common.

4. A comprehensive approach to the study of misutilization might well be guided by the same six honest serving men who framed my initial discussion of this paper of utilization.

1. What is misused?
2. Who misuses evaluation?
3. When is evaluation misused?
4. How is it misused?
5. Where is misused?
6. Why is it misused?

What better way to approach dishonesty and misuse than by mobilizing these six honest serving men in the service of appropriate evaluation utilization?

References

Alkin, Marvin C., Richard Daillak, and Peter White. Using evaluations: Does evaluation make a difference? Beverly Hills, CA: Sage, 1979.

Alkin, Marvin C. and Fred Ellett. "Evaluation Models," International Encyclopedia of Education. New York: Pergamon Press, 1984.

Alkin, Marvin C. and Lewis C. Solmon (Eds.) The Costs of Evaluation. Beverly Hills, CA: Sage, 1983.

Argyris, Chris. Interpersonal competence and organizational effectiveness. Homewood, IL: Irwin, 1972.

Argyris, Chris. Theory in practice: Increasing professional effectiveness. San Francisco: Jossey-Bass, 1974.

Argyris, Chris. Increasing leadership effectiveness. New York: John Wiley, 1976.

Azumi, Koya, and Jerald Hage (Eds.). Organizational systems. Lexington, MA: D.C. Heath, 1972.

Bennis, Warren. Changing organizations. New York: McGraw-Hill, 1966

Bennis, Warren, K. Berne, R. Chen, and K. Corey (Eds.) The Planning of change. New York: Holt, Rinehart & Winston, 1976.

Bernstein, Ilene, and Howard Freeman. Academic and entrepreneurial research. New York: Russell Sage, 1975.

Braskamp, Larry, and R.D. Brown (Eds.) Utilization of evaluative information. San Francisco: Jossey-Bass, 1980.

Burry, James. "Synthesis of the Evaluation Use Literature" Evaluation Productivity Project, Center for the Study of Evaluation, UCLA, 1984.

Chelimsky, Eleanor, "Improving the Cost Effectiveness of Evaluation, pp. 149-170 in Alkin and Solmon (1983).

Dickey, Barbara. Utilization of evaluations of small-scale educational projects. Educational Evaluation and Policy Analysis, 1981, 2(6), 65-77.

Fairweather, G.W., D. Sanders, and L. Tornatzky. Creating change in mental health organizations. New York: Pergamon, 1974.

Gephart, William J. Watercolor painting. Pp. 247-272 in Nick L. Smith (Ed.), Metaphors for evaluation. Beverly Hills, CA: Sage, 1981.

Guba, Egon G., and Yvonna S. Lincoln. Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches. San Francisco: Jossey-Bass, 1981.

Joint Committee on Standards fro Educational Evaluation. Standards for evaluations of education programs, projects, and materials. New York: McGraw-Hill, 1981.

Lynn, Laurence E. Designing public policies: A casework on the role of policy analysis. Santa Monica, CA: Goodyear, 1980.

Patton, Michael Q. Utilization-focused evaluation. Beverly Hills, CA: Sage, 1978.

Patton, Michael Q. Qualitative evaluation methods. Beverly Hills, CA: Sage, 1980.

Patton, Michael Q. Creative evaluation. Beverly Hills, CA: Sage, 1981.

Rossi, Peter H., Howard E. Freeman, and Sonia R. Wright. Evaluation: A systematic approach. Beverly Hills, CA: Sage, 1979.

Smith, Nick L. Studying evaluation assumptions. Evaluation Network Newsletter, 1980, Winter, 39-40.

Stevens, William F., and Louis G. Tornatsky. The dissemination of evaluation: An experiment. Evaluation review, 1980, 4(3), 339-354.

Thompson, Mark. Evaluation for decision in social programmers. Lexington, MA: D.C. Heth, 1975.
Tyler, Ralph W. Basic principles of curriculum and instruction. Chicago: University of Chicago Press, 1949.
Weiss, Carol H. (Ed.). Using social research in public policy making. Decision makers frames of reference for social science research. American Sociological Review, 1980, April, 302-313.
Worthen, Blaine R., and J.R. Sanders. Educational evaluation: Theory and practice. Worthington, OH: Charles A. Jones, 1973.